Understanding the Uses, Approaches and Applications of Sentiment Analysis

Peter Appiahene (peter.appiahene@uenr.edu.gh)
University of Energy and Natural Resources

Stephen Afrifa
Tianjin University

Emmanuel Akwa Kyei
University of Energy and Natural Resources

Peter Nimbe
University of Energy and Natural Resources

Research Article

Keywords: Sentiment Analysis (SA), Machine Learning (ML), Lexicon-based approach, Social Network Analysis (SNA), Natural Language Processing (NLP).

Posted Date: May 31st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1670421/v1

Abstract

Sentiment Analysis (SA) is a field of text mining research that is still evolving. SA is the algorithmic treatment of text's opinions, sentiments, and subjectivity to determine if a text contains negative, positive, or neutral feelings. We present a thorough introduction to sentiment analysis. This approach is deduced in a simple, intuitive manner, and implementation advice is provided. The many types, uses, challenges and techniques for sentiment analysis, as well as examples, are also explored in this study. The major goal of this survey is to provide a near-complete picture of SA techniques and related topics, compare sentiment analysis and social network analysis in the latter section of the paper, highlighting the distinctions and how they can both be used along with brief information.

1 Introduction

Sentiment analysis (also known as opinion mining) is a natural language processing (NLP) technique for determining the positive, negative, or neutral nature of data (MonkeyLearn, 2017). Sentiment analysis, according to Gupta (2018), is a type of text mining that finds and extracts subjective information from source material, allowing a company to determine the social sentiment of its brand, product, or service while monitoring online conversations. Sentiment analysis, often known as opinion mining, is an approach to natural language processing (NLP) that determines the emotional tone behind a body of text, according to TechTarget.com (2020). Both the terms SA and OM are interchangeable. They have the same meaning. However, some specialists believe that OM and SA have slightly different perspectives (Tsytsarau and Palpanas, 2012). Sentiment Analysis detects and analyzes the sentiment represented in a document, whereas Opinion Mining extracts and analyzes people's opinions on a subject. As a result, SA's purpose is to find people with strong beliefs, figure out what they're saying, and then characterize their polarity. Sentiment analysis assist firms in gaining a better understanding of the conversations and discussions that are taking place about them, as well as reacting and taking action in response. They can immediately detect any negative feelings voiced by customers and transform bad customer experiences into excellent ones.

Local government departments can use Facebook and Twitter comments to gauge public sentiment toward their department and the services they provide, and then use that information to improve services like parking and leisure facilities, local policing, and road conditions by listening to and analyzing comments. To gain a better understanding of their customers’ sentiments, businesses might compare their results to those of their competitors. They will be able to identify where they’re succeeding and where they may enhance their performance in contrast to the competition. With these great advantages it has, various obstacles must be solved before sentiment analysis can become a more useful tool and perfect as an automated system. In this study we discuss the types, uses, tools, algorithms, pros and cons and the similarities between SA and social network Analysis. Because of the usefulness and the exposition gained by sentiment analysis, applications are appearing steadily in the current literature. Researchers Raza et al. (2019), Taboada et al. (2022) and Zhang et al. (2014) have worked on applying sentiments analysis, that is, both on the lexicon and/or machine learning approaches. Despite its obvious

importance on real world issues, it is a field where much attention must be given to help policymakers to make informed decisions. This paper attempts to provide an intuitive knowledge on Sentiment Analysis to energize the interest of researchers and what it entails. We provide a tutorial introduction to sentiment analysis, deriving the types it entails, issues related to implementation and illustrations. We also discuss applications of the methods, one for lexicon approach and the other machine learning approaches. Moreover, the sectors in which sentiment analysis is applicable are also demonstrated in this paper.

A sentiment analysis of product reviews is depicted in the Fig. 1 below.

2 Types Of Sentiment Analysis

Sentiment analysis is concerned with a text's polarity (positive, negative, or neutral), but it may also detect specific moods and emotions (angry, joyful, sad, etc.), urgency (urgent, not urgent), and even intents (interested v. not interested). Social media sites such as Twitter, Facebook, and YouTube are becoming a valuable source of data known as social data (Manguri et al., 2020). Sentiment analysis can help individual or an organization to make decisions for further instructions. Below are types of sentiment analysis which are used to express opinions on issues such as policies, tax, etc.

2.1 Fine-grained Sentiment Analysis

Graded Sentiment is used to understand ratings. The ratings are expressed based how satisfied a person shows interest and/or dissatisfaction towards a policy. Ratings can be grouped as follows;

- 5 – Positive
- 4 –Very Positive
- 3 – neutral
- 2 – Very negative
- 1 – Negative

In fined grained sentiment analysis, 5 is expressed as highly positive (a person is more interested), 4 means a positive sentiment, 3 represents neutral (a person is neither against nor satisfied, and or have not understand what the content at hand means), 2 is not satisfied (very negative) with the content and 1 represents a negative sentiment. A fined grained sentiment analysis was carried out by Wang et al. (2017) described a social media analytics engine that employs a social adaptive fuzzy similarity-based classification method to automatically classify text messages into sentiment categories (positive, negative, neutral and mixed), with the ability to identify their prevailing emotion categories (e.g., satisfaction, happiness, excitement, anger, sadness, and anxiety). Their paper also embedded within an end to-end social media analysis system that has the capabilities to collect, filter, classify, and analyze social media text data and display a descriptive and predictive analytics dashboard for a given concept. In a research by Munikar et al. (2019) used a promising deep learning model called BERT to solve the fine-grained sentiment classification task. Their experiments showed that their model outperforms other

popular models for this task without sophisticated architecture. They also demonstrated the effectiveness of transfer learning in natural language processing in the process.

2.2 Emotion detection

Emotion detection sentiment analysis deals with interpreting emotions like happiness, frustration, anger, and sadness. Emotion detection systems frequently employ lexicons, which are collections of words that express specific emotions. Robust machine learning (ML) algorithms are also used by some advanced classifiers. Researchers Sagum et al, (2021) experimented on emotion detection result to be used in sentiment analysis. The emotions that were included in their research are happiness, sadness, anger, and fear. Once emotion was detected the system will then use it to know the sentiment of the person on a particular movie. Their paper aimed to measure the accuracy in sentiment analysis enhanced by emotion detection and to know whether emotion detection plays a key role in reading sentiment analysis. Kusal et al. (2021) also developed an Ai based emotion detection on big textual data. They considered 827 Scopus and 83 Web of Science research papers from the years 2005–2020 for the analysis. Their qualitative review represented different emotion models, datasets, algorithms, and application domains of text-based emotion detection. Their quantitative bibliometric review of contributions presents research details such as publications, volume, co-authorship networks, citation analysis, and demographic research distribution. In the end, challenges and probable solutions were showcased, which can provide future research directions in the area.

2.3 Aspect-Based Sentiment Analysis

When analyzing text sentiments, you will usually want to determine the specific qualities or features people are referencing in a favorable, neutral, or negative light. Aspect-based analysis entails a more in-depth investigation. It assists you in determining which components of the conversation are being discussed. It takes into consideration the whole sentence or text. In a related work on aspect-based sentiment analysis, Alqaryouti et al. (2019) proposed an aspect-based sentiment analysis hybrid approach that integrates domain lexicons and rules to analyze the entities smart apps reviews. The proposed model aimed to extract the important aspects from the reviews and classify the corresponding sentiments. This approach adopted language processing techniques, rules, and lexicons to address several sentiment analyses challenges, and produce summarized results. According to the reported results, the aspect extraction accuracy improves significantly when the implicit aspects are considered. In another work, Hoang et al. (2019) showed the potential of using the contextual word representations from the pre-trained language model BERT, together with a fine-tuning method with additional generated text, in order to solve out-of-domain ABSA and outperform previous state-of-the-art results on SemEval-2015 (task 12, subtask 2) and SemEval-2016 (task 5). According to the researchers, no other existing work has been done on out-of-domain ABSA for aspect classification.

2.4 Intent Analysis

The intent analysis can assist you figure out whether a customer is looking to buy anything or is just looking around. If a customer is willing to make a purchase, you can monitor them and market to them.

You can save time and money by not advertising to customers who aren't ready to buy. Lye & The (2021) analyzed data on customer Feedback in the form of Net Promoter Score (NPS) with a text box and demonstrated a hybrid representation that resulted in the accuracy improvement of the sentiment classification task and predicting customer intent. Their datasets were first trained using Word2Vec with the previous dataset and then fit into the Random Forest classifier, tested as the best configuration to prevent overfitting. The hybrid representation was compared against the baseline sentiment polarity tool through few experiments; the results showed that the hybrid model has improved accuracy for the sentiment classification task. Lastly, they performed customer intent prediction by using the Power BI influencer module.

2.5 Multilingual Sentiment Analysis

Helps detect language in texts automatically with a language classifier, then train a custom sentiment analysis model to classify texts in the language of your choice. Coding experience is required since it is difficult. Atiqah et al. (2021) in a study provided a systematic literature review on multilingual sentiment analysis, which summarized the common languages supported in multilingual sentiment analysis, pre-processing techniques, existing sentiment analysis approaches, and evaluation models that have been used for multilingual sentiment analysis. By following the systematic literature review, their findings revealed, most of the models supported two languages, and English is seen as the most used language in sentiment analysis studies. In a related study, Dashtipour et al. (2016) presented a state-of-the-art review on multilingual sentiment analysis. More importantly, they compared their implementation of existing approaches on common data. Precision observed in their experiments was typically lower than the one reported by the original authors, which they attributed to the lack of detail in the original presentation of those approaches.

3 Uses Of Sentiment Analysis

At this point we are assured that; the reader has come to terms with the types of sentiments and their application fields in the real world. In this section we explain the uses of Sentiment Analysis solving real-world problems. The results of sentiment analysis assist firms in gaining a better understanding of the conversations and discussions that are taking place about them, as well as reacting and taking action in response. They can immediately detect any negative feelings voiced by customers and transform bad customer experiences into excellent ones. They may improve their products and services, as well as the marketing messages they put out, by listening to what their target audience or customers are saying. All of this translates to higher revenue and sales. Listening to and analyzing comments on Facebook and Twitter can help local government departments gauge public sentiment toward their department and the services they provide, and then using that information to improve services like parking and leisure facilities, local policing, and road conditions.

Universities can employ sentiment analysis to examine student opinions and comments gathered from surveys or internet sources such as social media. They can then use the information to identify and resolve any areas of student unhappiness, as well as to identify and expand on areas where students are

expressing good feelings. Businesses may compare their results to those of their competitors to have a better understanding of their customers' attitudes. They will be able to see where they're succeeding and where they may improve in comparison to the competition.

4 Sentiment Analysis Tools

Sentiment Analysis may clearly be utilized by governments, agencies, institutions, and individuals to make educated decisions. There are tools which are used to implement Sentiment Analysis and are termed as sentiment analysis tools. Sentiment Analysis Tools are Artificial intelligence (AI) technologies that analyses text data to help you quickly determine how people feel about a policy, brand, product, or service (Maharani and Effendy, 2022). These AI technologies help businesses and organizations to make decisions beforehand. Monkey Learn, Awario, Brandwatch, Talkwalker, Lexalytics, and Hootsuite Insights are tools which can be used in sentiment analysis.

5 Pros And Cons Of Sentiment Analysis

In every application or technologies which are used for analyses and or other matters faces some shortcomings, however, we believe these are minimal to the operations of the tools, inasmuch as they also have their advantages. Sentiments Analysis faces the aforementioned as there is with any other tools. The pros and cons with regards to sentiment analysis are discussed in sections 5.1 and 5.2.

5.1 The Pros Associated with Sentiment Analysis

From more empathic service for each client to smarter chatbots to a better knowledge of your support team's and brand's overall performance, sentiment analysis has several benefits. It benefits everyone from businesses to governments to people. It helps businesses develop brands, provide upselling opportunities, agent monitoring to provide a clear perspective of customer happiness, and train chat bots to recognize and respond to client moods, all of which contribute to a stronger brand and a competitive edge. It assists government and leaders in understanding the views voiced by their constituents regarding projects and policies.

5.2 The Cons Associated with Sentiment Analysis

Despite the potential benefits, there are some disadvantages to employing automatic analysis, such as the difficulty in applying it due to the ambiguity of natural language and the peculiarities of the data presented. This can be seen in the examination of tweets, which are usually accompanied by hashtags, emoticons, and links, making it difficult to understand the communicated mood. Automatic processes are also necessary, which necessitate large datasets of annotated posts or lexical databases containing emotional phrases connected with sentiment scores. Furthermore, computer systems have difficulty comprehending sarcasm and irony, negations, jokes, and exaggerations, which a human would perceive, and failing to detect them might result in distorted results. For sentiment analysis purposes, the word "disappointed" may be categorized as a negative term, but it should be classified as positive within the

statement "I wasn't disappointed." Another important element is that the analyses are written in English, which is a disadvantage for other languages.

6 Sources Of Data For Sentiment Analysis

Textual data is frequently used in sentiment analysis. These data include product and consumer reviews, information from social media sites like as Facebook, Twitter, and YouTube, and much more. Twitter as a microblogging platform provides textual data for sentiment analysis which are used to express opinions on issues, products, or feelings. Amazon also provide reviews on products which sentiment analysis can be applied on them, movie websites as well as websites of institutions can be applied on sentiment analysis. These are aimed at giving feedback to potential investors and stakeholders to make decisions.

7 Sentiment Algorithms

Artificial intelligence has ushered in new development opportunities alongside the rapid advancement of science and technology. For sentiment analysis, there are two basic methodologies: lexicon-based and machine-learning-based approaches. Multidisciplinary theoretical knowledge, such as statistics and algorithm complexity, is incorporated into machine technology based on computer technology, which increases the functional features of artificial intelligence (Jin, 2020). Machine Learning is concerned with the use of data and algorithms to mimic the way humans learn, with the goal of gradually improving accuracy. The sentiment lexicon, which contains information on which words and phrases are positive and which are negative, is used in the Lexicon-Based Approach (Bonta et al., 2019). A sentiment lexicon is a collection of lexical properties that are classified as positive or negative depending on their semantic orientation. Figure 2 below shows a representation of the sentiment algorithms.

The subsequent sections below will describe the various types of algorithms used for sentiment analysis. These will serve as a guideline for the reader to appreciate the algorithms used for sentiment analysis.

7.1 Rule or Lexicon Approach

To determine sentiment, this method relies on manually generated data classification rules. To generate a score. This method represents polarity and sentiment strength by using dictionaries of words with positive or negative values. Expressions can also be used to offer additional functionality. By building even smarter rules, rule-based sentiment analysis algorithms can be adjusted based on context. A brief description of the techniques used are described below.

7.1.1 Sentiment Lexicon Generation Using Corpus-Based Techniques

This method uses co-occurrence data or syntactic patterns from text corpora, as well as a set of positive and negative seed words. To aid with polarity assignment, information in the context around the target

term might be used. Using a corpus in the target domain, this strategy is also widely used to modify a domain independent sentiment lexicon into a new domain-specific vocabulary (Darwich et al., 2019).

7.1.2 Dictionary-Based Sentiment Analysis

The concept underlying dictionary-based sentiment analysis is simple: a text's sentiment is defined by the sentiment of its words (Shi, 2019). This sort of analysis determines sentiments based on dictionaries developed by humans, with each word connected with its own sentiment. Of course, those dictionaries do not contain every word in the English (or other) language(s), but it is also true that the vast majority of English words are neutral, with only a few carrying distinct emotive connotations (Shi, 2019).

7.1.3 Related Works Based on Lexicon Approach

Yadav et al. (2018) focused on a lexicon-based strategy that relies on an external dictionary in their investigation. Their goal was to divide the tweets into two categories: positive and negative. They calculated the score by extracting the semantics from the tweets. This score aids in the categorization of tweets into positive or negative categories. They used the R programming language in this sentiment analysis project. In a related study, Bonta et al. (2019) employed the NLTK, Text blob, and VADER Sentiment analysis tools to categorize movie reviews acquired from the Cornell University website www.rottentomatoes.com, and compared these methods to discover the most efficient one for sentiment classification. The results of this study's experiments show that VADER outperforms the Text blob.

7.1.4 Disadvantages of the Lexicon Approach

The lexicon approach of sentiment algorithm comes with their disadvantages. The lexicon approach simply looks at occurrences and ignores how words are integrated in a phrase. Moreover, it is simple to execute, but the model has a long-term cost because it necessitates ongoing maintenance to ensure consistent and improved results.

8 Sentiment Analysis And Machine Learning Algorithms

A range of strategies and sophisticated algorithms are utilized to command and train machines to perform sentiment analysis. Each has its own set of benefits and drawbacks. However, when used in unison, they can produce incredible results. The subsections in section 8 below address some of the most often utilized algorithms.

8.1 Naive Bayes

Naive Bayes is a relatively simple collection of probabilistic algorithms for sentiment analysis categorization that assigns a probability that a certain word or phrase should be viewed positive or negative. Naive Bayes gets more accurate in terms of prediction when techniques like lemmatization, stop-word removal, and TF-IDF are utilized. This is how Bayes' theorem works in practice. The probability of B being true if A is true divided by the probability of B being true equals the chance of A being true if B is true:

$$\mathrm{P(A \mid B)} = \frac{P(B \mid A)\,P(A)}{P(B)} \quad (1)$$

According to Bayes' theorem, the probability of A if B is true is equal to the likelihood of B if A is true times the probability of A being true, divided by the probability of B being true.

8.2 Linear Regression

Linear regression is a statistical process that predicts a Y result using X data. The data sets are evaluated using machine learning to discover if there is a correlation. To predict future relationships, the relationships are represented on an X/Y axis with a straight line linking them. Linear regression is used to determine the relationship between the X input (words and phrases) and the Y output (polarity). This will determine where words and phrases fall on a polarity scale ranging from "very positive" to "extremely negative," and all points in between.

8.3 Support Vector Machines (SVM)

A support vector machine is a supervised machine learning model that is similar to linear regression but more advanced. SVM, which goes beyond X/Y prediction in our emotion polarity model, uses algorithms to learn and classify text. When employing SVM, the more sophisticated the data, the more accurate the predictor will become. SVM allows for more accurate machine learning because it is multidimensional.

8.4 Deep Learning

Deep learning is a subfield of machine learning that uses "artificial neural networks" to calculate data in the same way that the human brain does. Deep learning is a type of machine learning that is structured in a hierarchy. It's multi-level, in other words, and allows a machine to 'chain' together a number of human-created procedures. Deep learning is able to solve complicated problems in the same manner people do by allowing many algorithms to be applied sequentially while proceeding from step to step.

8.5 Related Works Based on Machine Learning Approach

Machine Learning can also be used to analyze sentiment and researchers have taken a clue from that angle. In a related study by Raza et al. (2019) carried out a scientific text sentiment by using an existing constructed annotated corpus. There were 8736 citation sentences in the corpus. To clean the data corpus, they used various data normalization algorithms to eliminate noise from the data. They created a system that implemented six different machine learning algorithms, including Nave-Bayes (NB), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), K-Nearest Neighbor (KNN), and Random Forest (RF), to conduct classification on their data set. Machova and Mach (2022) also conducted research to address the challenge of spotting suspicious reviewers in online social networking discussions. They focused on trolls, a specific type of suspicious author. They employed machine learning approaches to create detection models to distinguish a troll reviewer from a regular reviewer, as well as sentiment analysis methods to discern the sentiment typical of troll remarks. In a study on feelings using Machine Learning, Kawade and Oza (2017) extracted tweets regarding the Uri incident and found emotions and polarity of tweets. Text mining techniques were utilized to extract emotions and

polarity from tweets. Approximately 5000 tweets are recoded and pre-processed as part of their research to build a collection of commonly occurring words.

8.6 Hybrid Approach of Sentiment Analysis

Hybrid sentiment analysis models are the most modern, efficient, and widely used approach to sentiment analysis. If you have well-designed hybrid systems, you can reap the benefits of both automatic and rule-based systems. Hybrid models can combine machine learning capabilities with customizable flexibility. In a paper, researchers Ray and Chakrabarti (2022) proposed a deep learning approach for extracting aspects from text and analyzing user sentiment related to the aspect. Each component of the opinionated phrases was tagged using a seven-layer deep convolutional neural network (CNN). To increase the performance of the aspect extraction and sentiment scoring methods, they merged a deep learning approach with a series of rule-based approaches. They also compared our proposed method to some of the state-of-the-art methods in order to improve the existing rule-based approach of aspect extraction by aspect categorization with a preset set of aspect categories. Nguyen et al. (2018) conducted a comparison of text sentiment classification models employing word frequency inverse document frequency vectorization in both supervised machine learning and lexicon-based techniques in their study. Logistic Regression (LR), Support Vector Machine (SVM), and Gradient Boosting are the three machine learning algorithms used.

9 Social Network Analysis And Sentiment Analysis

Hundreds of millions of Internet users across the world may create and consume content through online social networks. They give users unprecedented access to a massive database of information. Even still, the raw data generated by these networks' users is a deluge of ideas, information, opinions, and so on. People are connected through social media platforms such as Facebook, Twitter, WhatsApp, and YouTube (Thangaraj and Amutha, 2017). SNA is a set of methodologies and tools that can be used to investigate the relationships, interactions, and communications in a social network (Saqr and Alamro, 2019). When connecting people through texts, images, audios and videos, it is made through some perfect representation by using the Graph (G = V, E). Here v stands for vertices and E stands for edges (Baecchi et al., 2016). Totally all the pairs connected and interconnected by objects. Each object is used and linked for the communication. Through the communication messages passes and relationship created and maintained by the social media. The links in a relationship can encode any type of relationship, such as familial, friendship, professional, or organizational (Liao et al., 2016). A useful way for consistently monitoring interactions in an online setting can be done using social network analysis. When individuals post, comment, or converse on sites like social media, sentiment analysis is used to assess whether the comments and concerns shared are positive, negative, or neutral.

9.1 Related Works Based on Social Network Analysis and Sentiment Analysis

Grandjean (2017) showed that sentiment analysis can be done via Twitter, a popular social media platform among this "community of practice." The depiction of the "who's following who?" graph, based on a network analysis of 2,500 people identified as participants of this movement, allows the researchers to highlight the structure of the network's relationships and identify users whose position is unique. They show, in particular, that linguistic groups are important determinants in explaining clustering inside a network with characteristics that resemble a tiny world. Fornacciari et al. (2016) offered a possible integrated technique of Social Network Analysis and Sentiment Analysis in another study. They attempted to assign a mood to the nodes of the graphs depicting social ties in particular, in the hopes of highlighting potential linkages. The idea is that the network topology can contextualize and then partially unmask some incorrect Sentiment Analysis results; on the other hand, the polarity of the feelings on the network can highlight the role of semantic connections in the hierarchy of the communities that are present in the network.

10. Application Of Sentiment

Sentiment Analysis can be applied in several sectors of the society to help stakeholders and policymakers to make informed decisions. From this research, we have demonstrated that, sentiment analysis can be applied using the Lexicon and/or Machine Learning approach. This paper has shown that many researchers have applied the two approaches, and or using the hybrid approach (a combination of the two approaches). For the purposes of this study, we applied Sentiment Analysis on University Teachers Association of Ghana (UTAG) and Lesbian, Gay, Bisexual, Transgender, and Queer (LGBTQ+) issues around the world to ascertain the reactions of the populace both on mainstream and/or social media.

10.1 The University Teachers Association of Ghana (UTAG)

The microblogging platform, Twitter, was used to collect the data. The data on University Teachers Association of Ghana (UTAG) was collected between August 8, 2021 to February 15, 2022. A total of 23,208 tweets were collected. We performed sentiment analysis on the dataset using the Lexicon approach to analyze the feelings of the general public of the ongoing strike action by the association.

Table 1 represents sample tweets that were collected based on the UTAG dataset.

Table 1
Sample tweets on UTAG strike

'See how @davido just got in shape after just few weeks.\n\nYvonne Nelson | UTAG | Tyga | #AshewoSpace | violence https://t.co/cVhaOmSBNx'
'How can I earn small moneyðŸ’µ for pocket as a studentðŸ¤”ðŸ¤”? Take gambling out.\n#UTAG\n#Tyga\n#AshewoSpace'
na obiaa ehu nu stands. #FadiBanku \n| UTAG | Sampson| Violence | Yvonne Nelson |Tyga | https://t.co/WPq6LwG5zy'
'Vals day promo till 12amâ€¦ 50ghc to join vip group dm and register till 12am\n\n#AshewoSpace UTAG Yvonne Nelson Tyga https://t.co/k554ufZnVj'
'UTAG dey strike no see what you people dey

The dataset was subjected to sentiment analysis in order to determine the impact of the union's nationwide strike. Figure 3 below shows the sentiments expressed by the general public on the UTAG strike.

From the analysis we can say that the general public were neutral on the issue of the UTAG strike. The mainstream and social media comments showed that, the general public were concerned by the demands by the teacher unions, that is, requesting for better conditions of service. The word count in Fig. 4 shows that, the president, Nana Addo Dankwa Akuffo-Addo was mentioned more inasmuch as requesting a consideration of student’s plight.

10.2 The LGBTQ + Dataset

The issue of LGBTQ + gained public discussion in Ghana when queer Ghanaians opened their office space in Ghana January 01, 2021 to provide a safe haven for their members. This opening ceremony saw in attendance the Denmark, and Australian Embassies in Ghana, as well as high and international figures in the country. The Ghanaian people shared their sentiments on the issue. In this study, we collected dataset on LGBTQ + around the globe to apply sentiment analysis on it. The data was collected between January 1, 2021 to February 15, 2022. A total of 726,998 tweets were collected. Table 2 below shows sample tweets on the LGBTQ + issues around the globe.

Table 2
Sample tweets on LGBTQ+

'Trwa przygotowanie do akceptacji LGBT w KoÅ›ciele? Reforma Kongregacji Nauki Wiary budzi niepokÃ³j
'CÃ³mo a Marvel le da miedo poner las dos bodas LGBT https://t.co/2UQ0JIVF84'
this movie is so lgbt i'm so fucking funny the scarecrow tin man and cowardly lion are all queer coded and I have absolutely no regrets saying this

The Fig. 5 shows the sentiments expressed on LGBTQ + issues around the world.

Figure 6 shows the word count on the sentiments expressed by the people.

From the above analysis we can say people around the globe share neutral sentiments when it comes to LGBTQ + issues, however, the positive sentiments are quite encouraging based on this study.

The researchers of this paper believe that sentiment analysis can be applied in many areas of the economy which can help governments, institutions, and organizations to make decisions on policies, products, or tax.

10.2 Limitations

When it comes to sentiment analysis, we have shown that it can be applied based on the lexicon and machine learning approaches. In spite of this, it comes with their challenges. The lexicon approach is language biased, that is, it favors the English language much than the other languages.

11 Conclusion

Sentiment analysis is a wonderful technique to figure out what the general public thinks about a company, product, policy, opinions, etc. It does, however, come with its own set of obstacles and restrictions, which may be overcome if properly utilized. It is sometimes difficult to identify the tone of comments, especially when irony and sarcasm are present. Furthermore, some algorithms are complex and may not yield particularly useful results. However, sentiment analysis is a great approach to collect unbiased feedback from clients on a variety of topics. It can assist businesses in a variety of ways, particularly when it comes to marketing and advertising as well as market research. In this paper we have demonstrated that, sentiment analysis can be used in many disciplines of the economy. The reader is abreast with the approaches involved in sentiments analysis. A practical application of sentiments is applied on datasets of University Teachers Association of Ghana (UTAG) and LGBTQ + to show the expressions of the people. We also demonstrated how a hybrid strategy combining lexicon and machine learning may be utilized to anticipate a reasonable result in sentiment analysis. These will help to make informed decisions and increase productivity where necessary.

11.1 Future Works

The lexicon approach to sentiment analysis is language biased. We are much more concerned about getting a corpus which will include the local languages in Ghana, with priority given to Akan language (widely spoken by the Ghanaian population). In future studies, we can use the hybrid approach to analyze the sentiments on the available data. The hybrid approach is extensively utilized approach to sentiment analysis. Due to language biasness, much of the data on UTAG were not captured to be analyzed. We strongly believe that, when corpus which contains the local languages is considered, it will improve the sentiment analysis performed on the data. Moreover, in a future study, we will combine sentiment analysis with machine learning or deep learning models to build a decision generation system on the data available.

Declarations

Availability of Data and Materials

The dataset generated and/or analyzed during the current study are available in the figshare repository through the link UTAG.csv (figshare.com) and LGBTQ.csv (figshare.com) for UTAG and LGBTQ+ respectively for free.

Competing Interest

The authors have no relevant financial or non-financial interests to disclose

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Peter Appiahene, Stephen Afrifa, Emmanuel Akwa Kyei and Peter Nimbe. The first draft of the manuscript was written by Peter Appiahene and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to express our profound gratitude to the microblogging site, Twitter, for making the data available to be used for our analysis in this current work.

Conflict of Interest

The authors declare that there is no conflict of interest in this paper.

References

1. Alqaryouti, O. *et al.* (2019) 'Aspect-based sentiment analysis using smart government review data', Applied Computing and Informatics. doi: 10.1016/j.aci.2019.11.003.
2. Atiqah, N. *et al.* (2021) 'Multilingual Sentiment Analysis: A Systematic Literature Review Multilingual Sentiment Analysis : A Systematic Literature Review', (September). doi: 10.47836/pjst.29.1.25.
3. Baecchi, C. *et al.* (2016) 'A multimodal feature learning approach for sentiment analysis of social network multimedia', Multimedia Tools and Applications, 75(5), pp. 2507–2525. doi: 10.1007/s11042-015-2646-x.
4. Bonta, V., Kumaresh, N. and Janardhan, N. (2019) 'A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis', Asian Journal of Computer Science and Technology, 8(S2),

pp. 1–6. doi: 10.51983/ajcst-2019.8.s2.2037.

5. Darwich, M. *et al.* (2019) 'Corpus-Based Techniques for Sentiment Lexicon Generation: A Review', Journal of Digital Information Management, 17(5), p. 296. doi: 10.6025/jdim/2019/17/5/296-305.
6. Dashtipour, K. *et al.* (2016) 'Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques', Cognitive Computation. Springer US, 8(4), pp. 757–771. doi: 10.1007/s12559-016-9415-7.
7. Fornacciari, P., Mordonini, M. and Tomauiolo, M. (2016) 'Social Network and Sentiment Analysis on Twitter: Towards a Combined Approach'.
8. Grandjean, M. (2017) 'A social network analysis of Twitter: Mapping the digital humanities community', Cogent Arts & Humanities. Cogent, 115(1). doi: 10.1080/23311983.2016.1171458.
9. Gupta, S. (2018) 'Sentiment Analysis: Concept, Analysis and Applications | by Shashank Gupta | Towards Data Science', *datascience*, pp. 1-undefined.
10. H. Manguri, K., N. Ramadhan, R. and R. Mohammed Amin, P. (2020) 'Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks', Kurdistan Journal of Applied Research, pp. 54–65. doi: 10.24017/covid.8.
11. Hoang, M., Bihorac, O. A. and Rouces, J. (2019) 'Aspect-Based Sentiment Analysis using BERT', *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, pp. 187–196. Available at: https://www.aclweb.org/anthology/W19-6120.
12. Jin, W. (2020) 'Research on Machine Learning and Its Algorithms and Development', *Journal of Physics: Conference Series*, 1544(1). doi: 10.1088/1742-6596/1544/1/012003.
13. Kawade, D. R. and Oza, K. S. (2017) 'Sentiment Analysis: Machine Learning Approach', (June). doi: 10.21817/ijet/2017/v9i3/1709030151.
14. Kusal, S. *et al.* (2021) 'AI Based Emotion Detection for Textual Big Data: Techniques and Contribution'.
15. Liao, C. *et al.* (2016) 'A hybrid epidemic model for deindividuation and antinormative behavior in online social networks', Social Network Analysis and Mining, 6. doi: 10.1007/s13278-016-0321-5.
16. Lye, S. H. and Teh, P. L. (2021) 'Customer Intent Prediction using Sentiment Analysis Techniques'. IEEE, pp. 185–190.
17. Machova, K. and Mach, M. (2022) 'Comparison of Machine Learning and Sentiment Analysis in Detection of Suspicious Online Reviewers on Different Type of Data'.
18. Maharani, W. and Effendy, V. (2022) 'Big five personality prediction based in Indonesian tweets using machine learning methods', International Journal of Electrical and Computer Engineering, 12(2), pp. 1973–1981. doi: 10.11591/ijece.v12i2.pp1973-1981.
19. Medhat, W., Hassan, A. and Korashy, H. (2014) 'Sentiment analysis algorithms and applications: A survey', Ain Shams Engineering Journal. Faculty of Engineering, Ain Shams University, 5(4), pp. 1093–1113. doi: 10.1016/j.asej.2014.04.011.
20. MonkeyLearn (2017) *Sentiment Analysis Guide*, *MonkeyLearn*.

21. Munikar, M., Shakya, S. and Shrestha, A. (2019) ‘Fine-grained Sentiment Classification using BERT’, *International Conference on Artificial Intelligence for Transforming Business and Society, AITB 2019*, pp. 2–5. doi: 10.1109/AITB48515.2019.8947435.
22. Nguyen, H., Al, R. and Academy, K. (2018) ‘Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches’, 1(4).
23. Ray, P. and Chakrabarti, A. (2022) ‘A Mixed approach of Deep Learning method and Rule-Based method to improve Aspect Level Sentiment Analysis’, Applied Computing and Informatics, 18(1–2), pp. 163–178. doi: 10.1016/j.aci.2019.02.002.
24. Raza, H. *et al.* (2019) ‘Scientific Text Sentiment Analysis using Machine Learning Techniques’, 10(12), pp. 157–165.
25. Sagum, R. A. (2021) ‘An Application of Emotion Detection in Sentiment Analysis on Movie Reviews’, Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(3), pp. 5468–5474. doi: 10.17762/turcomat.v12i3.2204.
26. Saqr, M. and Alamro, A. (2019) ‘The role of social network analysis as a learning analytics tool in online problem based learning’, *BMC Medical Education*. BMC Medical Education, 19(1), pp. 1–11. doi: 10.1186/s12909-019-1599-6.
27. Shi, F. (2019) ‘Learn About Sentiment Analysis With Supervised Learning in Python With Data From the Economic News Article Tone Dataset (2016)’, *Learn About Sentiment Analysis With Supervised Learning in Python With Data From the Economic News Article Tone Dataset (2016)*, (2016). doi: 10.4135/9781526499608.
28. Taboada, M., Brooke, J. and Voll, K. (2022) ‘Lexicon-Based Methods for Sentiment Analysis’, (September 2010).
29. TechTarget.com (2020) ‘What is sentiment analysis (opinion mining)? - Definition from WhatIs.com’, *TechTarget*, pp. 2–3.
30. Thangaraj, M. and Amutha, S. (2017) ‘Similarity between Sentiment Analysis and Social Network Analysis’, International Journal of Scientific & Engineering Research, 8(4), pp. 1526–1532.
31. Tsytsarau, M. and Palpanas, T. (2012) ‘Survey on mining subjective data on the web’, Data Mining and Knowledge Discovery, 24(3), pp. 478–514. doi: 10.1007/s10618-011-0238-6.
32. Wang, Z. *et al.* (2017) ‘Fine-grained sentiment analysis of social media with emotion sensing’, *FTC 2016 - Proceedings of Future Technologies Conference*, (December), pp. 1361–1364. doi: 10.1109/FTC.2016.7821783.
33. Yadav, N. N. and D. (2018) ‘Lexicon-based approach to Sentiment Analysis of tweets using R language’, (April). doi: 10.1007/978-981-13-1810-8.
34. Zhang, H., Gan, W. and Jiang, B. (2014) ‘Machine Learning and Lexicon based Methods for Sentiment Classification: A Survey 1’. IEEE, pp. 262–265. doi: 10.1109/WISA.2014.55.

Figures

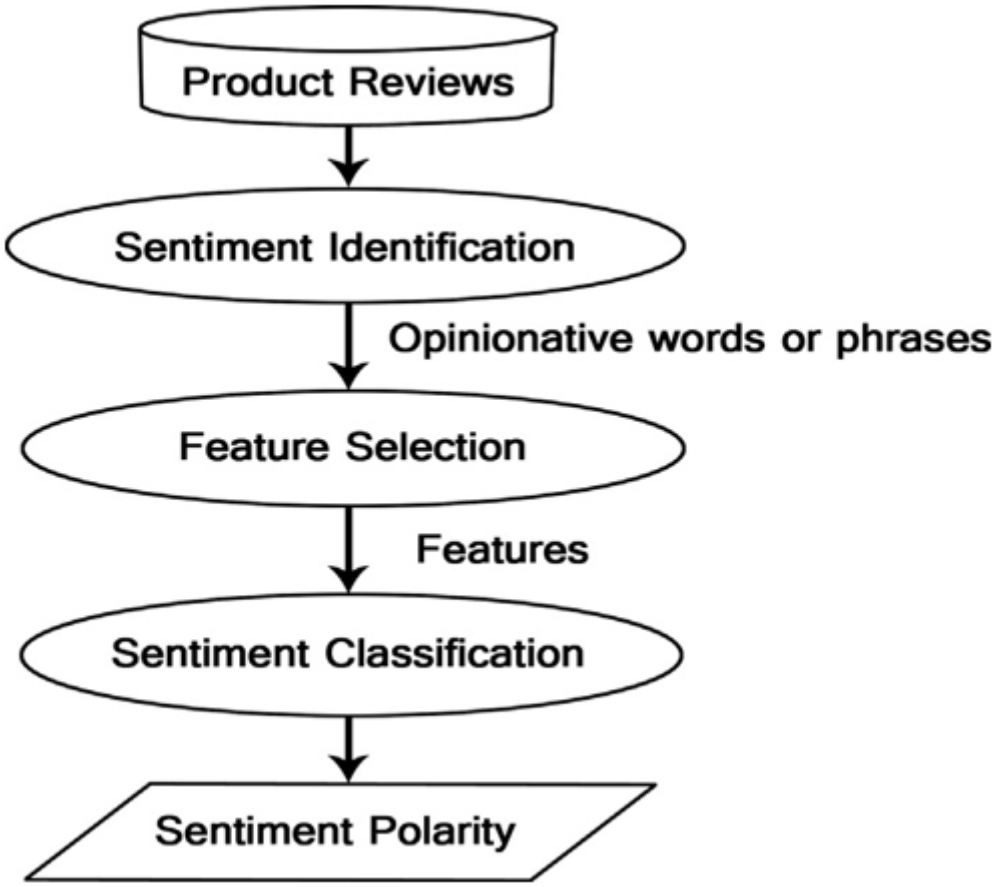

Figure 1

Sentiment Analysis product review (Source: Medhat et al., 2014)

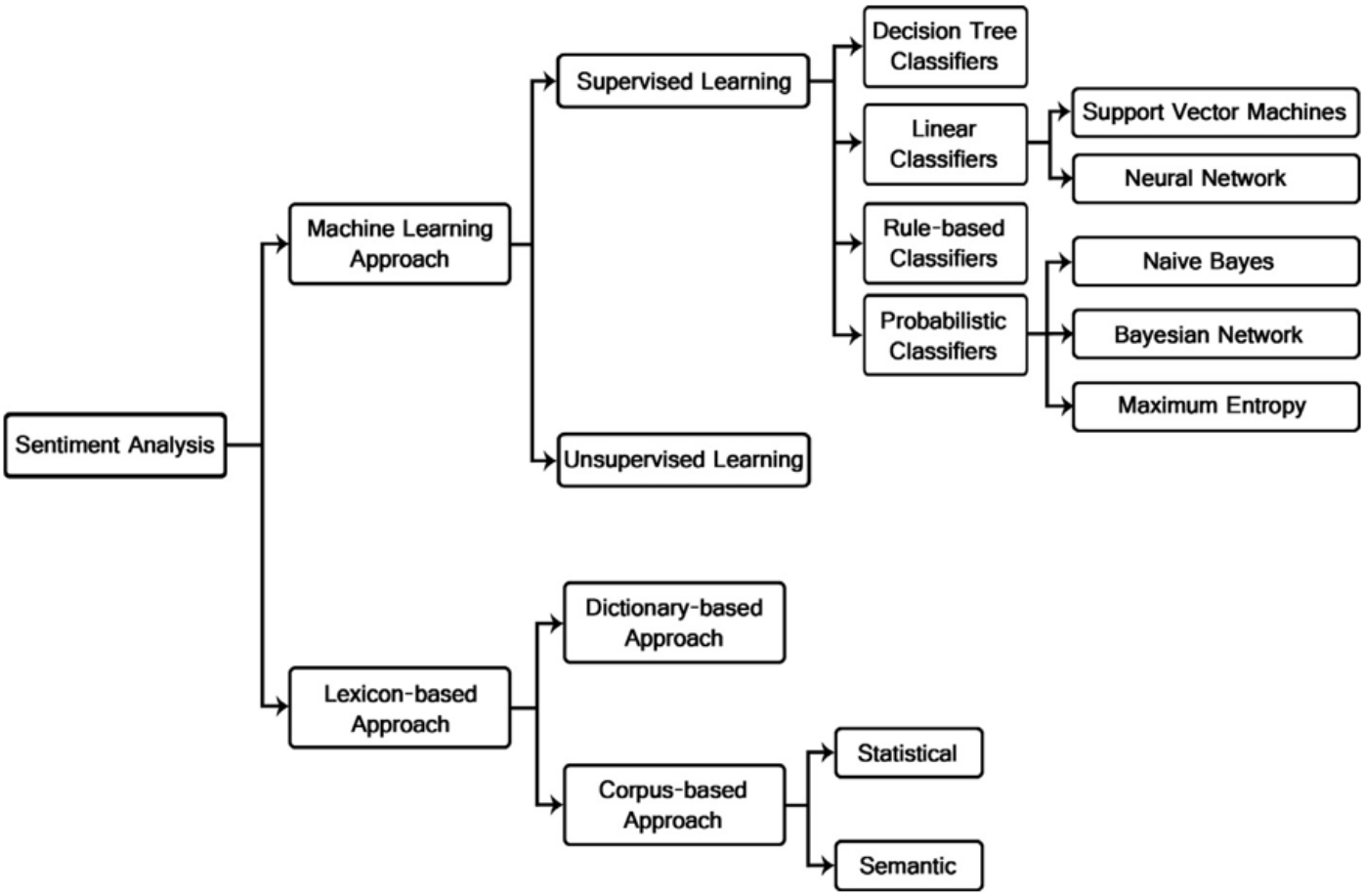

Figure 2

Sentiment Algorithms for Analysis (Source: Medhat et al., 2014)

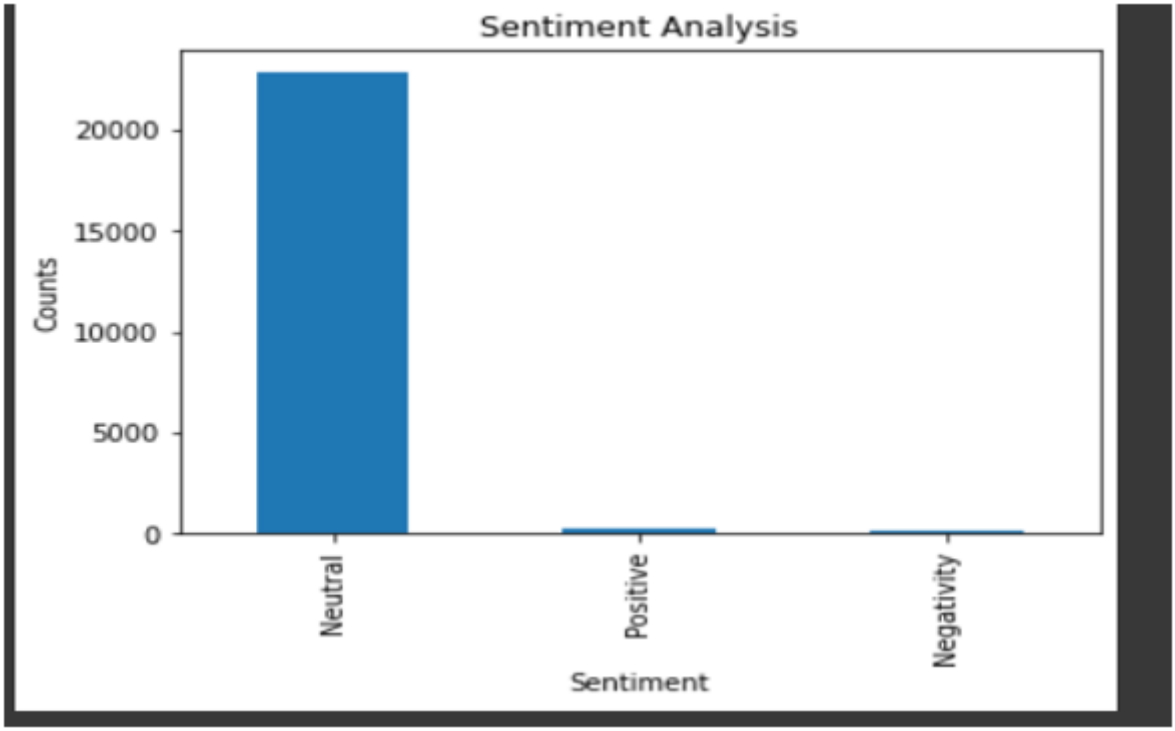

Figure 3

Sentiment analysis on UTAG dataset

Figure 4

Word count for the UTAG dataset

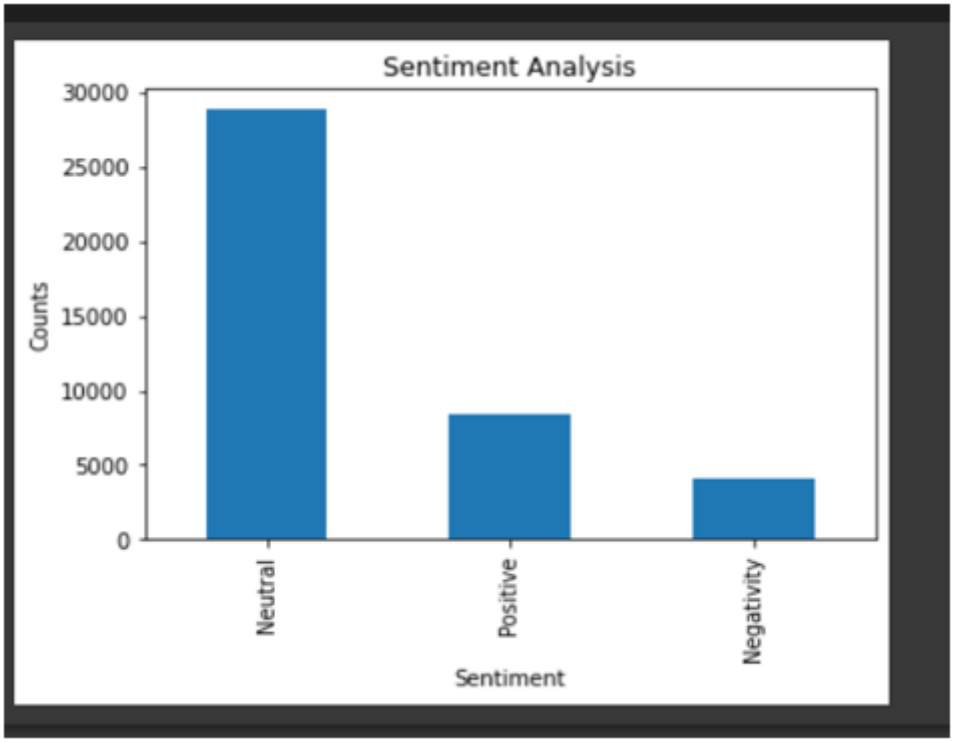

Figure 5

Sentiment analysis on LGBTQ+ in the world

Figure 6

Word count on LGBTQ+